DOCUMENT RESUME

ED 319 805 TM 015 111

AUTHOR Thayer, Jerome D.
TITLE Implementing Variable Selection Techniques in Regression.
PUB DATE Apr 90
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Computer Software; Decision Making; Methods Research; *Multiple Regression Analysis
IDENTIFIERS *Selection Tools; *Stepwise Regression; Variables

ABSTRACT

Variable selection techniques in stepwise regression analysis are discussed. In stepwise regression, variables are added or deleted from a model in sequence to produce a final "good" or "best" predictive model. Stepwise computer programs are discussed and four different variable selection strategies are described. These strategies include the forward method, backward method, forward stepwise method, and backward stepwise method. General and specific criticisms of the stepwise method in the literature are outlined, and misuses of the method are considered. Although most statisticians would agree that stepwise methods should not be used when an explanatory model is desired, it is common to see research articles where explanatory interpretations are given to a model that is called a "prediction" model. Stepwise methods should not be used to determine the number of variables in the final model. When model selection is being performed, the stepwise method can be helpful if: the initial choice of variables is conducted based on theory, the defaults are not used automatically, more than one run is done using different variable selection methods, and the final model is chosen through an intelligent process rather than automatically using the final model generated by the computer program. Three tables and outlines of the models described are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JEROME D. THAYER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED319805

Implementing Variable Selection Techniques in Regression

Jerome D. Thayer

Andrews University

BEST COPY AVAILABLE

TM015111

A paper presented as part of a symposium entitled "Variable Selection in Multiple Linear Regression" at the annual meeting of the American Educational Research Association, Boston, Massachusetts, April, 1990

Multiple regression is one of the most popular statistical techniques used in behavioral science research. There are three ways in which it is typically used:

1) Testing a full model, interpreting the model and each of its components.

2) Adding components to a model and interpreting the value of the increment.

3) Using a stepwise method in which variables are added or deleted from a model in sequence to come up with a final "good" or "best" predictive model.

This paper deals with the third of these methods, the stepwise method.

Defining the term "stepwise"

In considering the stepwise method it is necessary to contrast the stepwise method used as a computer program with the stepwise method used as a methodological procedure and to note the different ways in which the stepwise method can be used.

Many computer programs are called "stepwise" programs because they can be used to build models using a stepwise method with default or user-specified alternatives controlling factors of the selection process including the criteria for entering and removing variables.

Stepwise computer programs can be used in four ways:

1) The program selects a model automatically using only the default values.

2) The program selects a model automatically using some or all user-specified values in place of default values.

3) The researcher uses the output of the program to help in selecting a model.

4) The program is used to make specified incremental tests by adding one or more variables to other variables.

Methods one and two are, almost without exception, the methods used in journal articles that claim to be using the "stepwise method". However, few of them specify what statistical criteria are used for adding and removing variables. In most cases the default values are probably used (method one). Critics of the stepwise method usually criticize the use of stepwise programs in either of the two automatic ways listed (methods one and two).

Since method three uses the professional judgment of the researcher in the selection of the final model, this procedure will be suggested as the appropriate use of the stepwise method in this paper. Method four uses the stepwise computer program, but it is not a use of the stepwise method so will not be considered here.

The stepwise method as a procedure can be used to describe at least four different variable selection strategies.

1) Forward method

   The selection begins with no variables in the model and variables are added one at a time if they meet the statistical criterion for entering variables.

2) Backward method

The selection begins with all variables in the model and variables are removed one at a time if they meet the statistical criterion for removing variables.

3) Forward stepwise method

This is a variation of the forward method in which at each step, before any variable is added, variables already in the model are considered for removal if they meet the statistical criterion for removing variables.

4) Backward stepwise method

This is a variation of the backward method in which at each step, before any variable is removed, variables not in the model are considered for addition if they meet the statistical criterion for entering variables.

Usually in journal articles the method used is just called "stepwise" with no indication of which of the four methods or procedures is used. The method that is used in most cases is probably the forward stepwise method which is the default procedure for most stepwise computer programs.

Criticisms of the stepwise method

The stepwise method has been frequently criticized by methodologists (Davidson, 1988; Huberty, 1989; Thompson, 1989) and almost all authors of textbooks on multiple regression (i.e., Berensen et al., 1983; Chatterjee & Price, 1977; Cohen & Cohen, 1975; Draper & Smith, 1981; Freund & Minton, 1979; Gunst & Mason, 1980; Kleinbaum & Kupper, 1988; Morrison, 1983; Myers, 1986; Neter et al., 1983; Pedhazur, 1982; Wittink, 1988; Younger, 1979). The criticisms are both general and specific. Two examples of general criticisms are:

> Someone has characterized the user of stepwise regression as a person who checks his or her brain at the entrance of the computer center. (Wittink, 1988, p. 259)

> Stepwise regression is probably the most abused computerized statistical technique ever devised. If you think you need stepwise regression to solve a particular problem you have, it is almost certain that you do not. Professional statisticians rarely use automated stepwise regression. (Wilkinson, 1984, p. 196)

Critics of the stepwise method suggest the following considerations for selecting a subset of predictors for a prediction model:

1) Selection of variables for a regression model should not be an automatic or mechanical process.

2) No one method will consistently select the "best" model.

3) There is no one "best" model according to any common criterion such as the maximum R^2.

4) The stepwise method should not be used to build models for explanatory research.

5) The stepwise method has limited usefulness when predictors are highly correlated, if a key set of variables work in combination, or when suppression exists.

6) The order in which variables enter the model should not be used as an indicator of the value of the variable as a predictor.

If a stepwise method is used to select a model in the automatic way that is most commonly found in the literature, it is quite likely that:

1) Other models with the same number of predictors may very well have a larger R^2.

2) Smaller models may very well predict an equivalent R^2.

3) Variables not included in the model may be just as good or better predictors than some of the variables in the model.

4) The variables will probably not enter the model in order of their importance in the final model.

Misuses of the stepwise method

In spite of these criticisms and suggestions, there are still many research studies reported in the recent literature in which these guidelines are violated. Most of these studies have the following characteristics:

1) Models selected by the computer are called the "best" or "optimum" model for maximizing the explained variance (R^2) with the minimum number of predictors (k).

2) No description is given of the process by which the model was selected other than the term "the stepwise method was used". In most cases an automatic process was probably used.

3) Explanatory interpretations are made by defining "good" predictors as those in the model and "poor" predictors as those not in the model.

4) The interpretation of the model includes a ranking of the variables in the model in terms of importance based on order of entry.

5) No mention is made of the interrelationship of the variables in the description of the procedures used or in the interpretation of the final model selected.

Examples of misuses

Examples of these uses/misuses of stepwise regression found in the educational literature in 1988 and 1989 include:

1) We found the "most consistent variables that are most closely associated" with the criterion.

2) Variable A was picked as the "main predictor."

3) We wanted to find the "optimum equation."

4) The analysis yielded an "optimum predictor equation with as few predictors as possible."

5) "This allows the most consistent variables that are most closely associated with learning to be identified."

6) "The use of [variable A and variable B] as predictors revealed that [variable A] predicted [variable Y]. [Variable B] proved to lack significant predictive utility." In this article a table reported that the zero-order correlations between Y and variables A and B were .49 and .48. Variable B did not appear in the final model.

Should stepwise methods be used?

Although most, if not all statisticians would agree that stepwise methods should not be used when an explanatory model is desired, it is common to see research articles where explanatory interpretations are made to a model that is called a "prediction" model. Even if a predictive model is being selected, determining the value of each of the predictors in the prediction model requires more than what the stepwise method provides. Order of entry should not be used for this purpose. Stepwise methods should not be used to determine the number of variables in the final model. If multicollinearity exists in the data set, stepwise methods are especially suspect. In most cases, either the multicollinearity should be removed by removing variables, or other procedures should be used.

Since from the critics' point of view the stepwise methods are usually used in an inappropriate manner, the question then is whether the stepwise method should be a recommended technique for statistical analysis, and if so, how should it be used.

The objective of this paper is to consider the conditions under which variable selection procedures such as stepwise procedures can be used appropriately in educational research.

Value of using stepwise methods

The stepwise method is appropriate for situations in which a prediction model is desired, not an explanation model. In these situations, it is best used for exploratory analysis where little theory is available to guide in the selection of variables for the prediction model (Wittink, 1988).

Stepwise methods are very helpful if used properly when a subset of predictor variables is needed to be selected. A major advantage of stepwise methods is that by examining the output of each step of the model building process the researcher can see how each variable acts in different combinations which can be used to help the researcher to select the variables for the final model.

Observing the change in the partial correlations (and/or regression coefficients) as variables are added and deleted gives a feel for the variables that is difficult to get in any other way. If both forward and backward stepwise methods are used in conjunction with an all possible subsets program such as BMDP9R, a great variety of "good" models can be examined. An earlier study by Thayer (1986) showed that the backward stepwise and all possible subsets methods frequently gave different models than the forward stepwise method, in some cases with much higher R^2 values with the same or slightly more predictors.

The value of each potential predictor can be examined by comparing the zero-order correlations with the partial correlations at each step in the stepwise process. If the partial correlations remain high relative to the zero-order correlation, then the researcher can be confident of the stability of the variable in many prediction situations. If the partial correlations change markedly, then it will take some analysis to determine the dynamics involved, particularly noting which variables seem to be causing the changes.

How not to use stepwise methods

If stepwise methods are used the following procedures should be avoided:

1) Stepwise methods should not be used alone as the only procedure, especially if the researcher is looking for the "best" or "optimum" prediction model. An all possible subsets program such as BMDP9R should be used in conjunction with stepwise methods. It is also very desirable to use both the forward and backward stepwise methods to examine alternative models. When one method is used the temptation is great to use the model that the computer selects as the final model. The final model should be selected as a result of many considerations, not only the statistical criterion used by the stepwise program.

2) Stepwise methods should not be used automatically using the default values. The default values of F(or p)-to-enter and F(or p)-to-remove are seldom appropriate for good model selection. Whether the default values are used or not, they should be specified in the reporting of the results.

3) The p values given for the increments at each step should not be taken at "face" value. Huberty (1989) suggests that "the tail probabilities . . . should not be taken too seriously. And one should certainly not refer such probabilities to conventional significance levels to determine the 'significance' of an entered or removed response variable."

How to run stepwise programs

If a stepwise program is used to provide data to the researcher for model selection, the following suggestions are offered:

1) Reduce the number of variables to work with to a size that will allow you to do an "all possible subsets" (BMDP9R) run. If computer memory permits, do a backward stepwise run to find the best 27 (or number that can be run by an all possible subsets program). If there are too many variables to do a backward run, then do a forward run with a very low F-to-enter, forcing in all theoretically important variables to find the "best" 27 (or so).

2) Allow theoretically important variables (variables that have been shown or are hypothesized to be "causal" variables) to be entered first by forcing them in the model or allowing them to be eligible for entrance if they satisfy the statistical criterion for entering variables.

3) Set low F-to-enter or high p-to-enter values, such as F = 0.00-2.00 or p = .10-1.00 (Myers, 1986; Wittink, 1988). You should allow the computer to enter more variables (if forward) or delete more variables (if backward) then desired for a final model, in order to consider more variables than you will use in the selected model. The major advantage of this is to allow more combinations of variables to be considered in the final model.

4) Run both forward and backward stepwise and all possible subsets procedures in order to consider alternative models and to examine the performance of the variables in different models.

   The forward stepwise method frequently gives smaller models than the backward stepwise method and the researcher can observe changes occurring in the partial correlations (and/or regression coefficients) of variables which give a feel for the stability of the variable (Thayer, 1986).

The backward stepwise method has an advantage over the forward stepwise method because combinations of variables that work together but not singly are considered. The forward stepwise method will miss them (Thayer, 1986).

The all possible subsets method encourages examination of more then one model by providing statistics for many models of varying sizes.

5) Cross-validate alternative models suggested by the stepwise and all possible subsets runs. This can be done either by generating an equation from half of the data and cross-validating it on the other half, or by selecting another sample for the cross-validation.

6) Select the final model intelligently by using as many of the following criteria as possible:

Each variable in the model should contribute a meaningful amount to the total R^2 of the model (the incremental R squared of that variable in addition to the others in the model). With a large N the "best" model may be a smaller model than that suggested by considering only the p values of the variables.

The variables selected should as much as possible be theoretically meaningful variables.

The variables selected should as much as possible have partial correlations (and regression coefficients) which are relatively stable in the various steps or with different models. As variables are added or deleted in the stepwise process, if the sign of the partial correlation and regression coefficient for a predictor changes, that variable may not perform well in a cross-validation situation. If a partial correlation (and regression coefficient) becomes larger as the model increases in size, the variable should be studied closely to see whether there is some suppression or multicollinearity in the data that needs to be considered in the selection of the final model.

The variables selected should appear in many "good" models. Variables that only work in a few combinations would be unlikely to work well in a prediction model with new data.

The model should be one of the best models considered in terms of cross-validation.

How to report stepwise results

If stepwise procedures are used properly, many decisions must be made concerning how to run the stepwise program and how to select the variables for the final model. It is important that these decisions should be included in the final report.

The following procedures used should be reported:

1) F-to-enter/remove or p-to-enter/remove values used.

2) Stepping method used: forward, forward stepwise, backward, or backward stepwise.

3) Default or substitute values used.

4) Which alternative models were examined.

5) Results of stepwise methods compared to those of the all possible subsets method.

6) How subjective judgment (theory, etc.) was used in selecting variables for the model.

The following statistical results should be reported:

1) For each variable considered:

   Zero-order correlations and partial correlations with the dependent variable at each step.

2) For each variable selected:

   Why it was selected.

   The stability of its regression coefficient, b or ß, (or its partial correlation) in different models.

3) For each variable not selected:

   Why it was not selected.

   Whether the variable was a good predictor in other combinations of variables tested or a good predictor alone.

Summary

When model selection is being done, the stepwise method can be helpful if the initial choice of variables is chosen as much as possible using theory, the defaults are not used automatically, more than one run is done using different variable selection methods, and the final model is chosen through an intelligent process, not automatically using the final model generated by the computer program.

Example

Appendices A-C report computer printouts and models of three variable selection computer runs: forward stepwise, backward stepwise and all possible subsets, using the BMDP2R and BMDP9R computer programs on data set A6 from Gunst and Mason (1980, pp 355, 363). These printouts, summary tables and comments inserted where appropriate, illustrate most of the points presented in this paper.

To make interpretation easier, the BMDP2R listing in Appendix A is a combined forward and backward run but the variables entered and removed are in the same order as they were with runs done using the forward stepwise and backward stepwise methods using F-to-enter/remove values of 2.00/1.99.

Table 1 reports a summary of the models selected by the forward and backward stepwise methods described in more detail in Appendix A. With the forward stepwise method, variable 2 was the first variable entered. If the default F-to-enter value of 4.00 had been used, variable 2 would not have entered and the 0-predictor model would have been selected.

Using F-to-enter/remove values of 2.00/1.99, the automatic forward stepwise method selected a 2-predictor model with an R^2 of .1495 which was the same as the best 2-predictor model found using the all possible subsets method. If the forward stepwise method would have been allowed to continue adding variables with F's below 2.00, the larger models selected became progressively worse compared to those identified as the "best" of the same model size by the all possible subsets method.

Table 1

Summary of Models Selected By Forward and Backward Stepwise Methods

Predictors	Step	Method	Variables in the Model	R^2
2	2	Forward	2 4	.1495
2	26	Backward	2 13	.0903
4	4	Forward	2 4 8 10	.2108
4	24	Backward	2 6 13 15	.1783
5	5	Forward	2 4 8 10 12	.2428
5	23	Backward	2 6 12 13 15	.2289
6	6	Forward	2 4 8 9 10 12	.2649
6	22	Backward	2 5 6 12 13 15	.2966
7	7	Forward	2 4 5 8 9 10 12	.2963
7	21	Backward	2 5 6 9 12 13 15	.3472

The model selected by the backward stepwise method was a 7-predictor model which was the same model chosen by the all possible subsets as the best model of any size. If smaller models had been chosen with the backward stepwise method, they would have become progressively poorer than the "best" model of the same size selected by the all possible subsets method. Forward stepwise gave better small models while backward stepwise gave better large models.

Appendix B reports the printout of the BMDP9R all possible subsets run on the A6 data. One 4-predictor model, four 5-predictor models, at least ten 6-predictor models, and at least ten 7-predictor models had C_p values lower than the recommended minimum value ($k+1$ where k is the number of predictors in the model). The model with the lowest C_p value was a 7-predictor model. Although this model was identified as the "best" model by BMDP9R, all of the models with lower than minimum C_p values could be considered to be "good enough" models. The best 2, 4, 5, 6, and 7 predictor models along with all other models with acceptable C_p values with these model sizes are reported in Appendices A-C.

Table 2 compares the models selected by the all possible subsets method with those of the forward stepwise and backward stepwise methods. The three methods never gave the same models with 2, 4, 5, 6, or 7 predictors.

The models selected by the forward stepwise method were identified by the all possible subsets methods as the best 2-predictor model, the 2nd best 4-predictor model, the 5th best 5-predictor model and not in the top ten 6 or 7-predictor models. The models selected by the backward stepwise method were identified by the all possible subsets methods as the 4th best 2-predictor model, not in the top ten 4 or 5-predictor model, the 2nd best 6-predictor model, and the best 7-predictor model.

Variables 4, 8 and 10 were three of the first four variables entered in the forward stepwise method but they were also three of the first five removed in the backward stepwise method. Using the order of entry criterion for importance would indicate that 4, 8 and 10 were some of the best variables if you used the forward stepwise method or some of the worst variables if you used the backward stepwise method.

If the forward stepwise method would have been used to select the "best" model and order of entry was used to indicate importance (which should not be done), variable 2 would be called the "most

important" variable and variable 4 the "next most important". If better models had been used, such as those shown in Appendix C, and contribution to R^2 was used as the criterion for importance, variable 2 would have been the "most important" in every model, but variable 4 did not appear in any of the models.

Table 2

Models Selected By All Possible Subsets, Forward Stepwise and Backward Stepwise Methods

Predictors	Method	Variables in the Model	R^2	All Possible Ranking
2	All Possible Subsets	2 4	.1495	
2	Forward	2 4	.1495	1st
2	Backward	2 13	.0903	4th
4	All Possible Subsets	2 4 5 9	.2278	
4	Forward	2 4 8 10	.2108	2nd
4	Backward	2 6 13 15	.1783	Not in top 10
5	All Possible Subsets	2 5 9 12 13	.2699	
5	Forward	2 4 8 10 12	.2428	5th
5	Backward	2 6 12 13 15	.2289	Not in top 10
6	All Possible Subsets	2 3 5 12 13 15	.3000	
6	Forward	2 4 8 9 10 12	.2649	Not in top 10
6	Backward	2 5 6 12 13 15	.2966	2nd
7	All Possible Subsets	2 5 6 9 12 13 15	.3472	
7	Forward	2 4 5 8 9 10 12	.2963	Not in top 10
7	Backward	2 5 6 9 12 13 15	.3472	1st

Variable 5 was the second best variable in two of the four "best" models in Appendix C and Table 3 but the least important variable in the other two models. Since these models are good competitors for the "best" model as explained later, it can been seen that even using contribution to R^2 is likely to mislead in indicating the importance of variables, since it can be so heavily dependent on what other variables are in the model.

The listing of the partial correlations for each variable for each step gives an indication of the stability of the variables. Variable 2, the "most important" variable, is very stable, while other variables are shown to vary somewhat. None of the variables changes signs while in the model (indicated by as asterisk in the printout).

Disregarding theory in the selection of models, there were four models with 5, 6, and 7-predictors that appear to be worthy of selection as a "best" model are listed in Table 3. More complete information on the models is provided in Appendix C.

Variables 2, 5, 12, and 13 appear in all of these models, variable 15 in four of the models, variable 9 in three of the models, and variables 3 and 6 in only two of the models. In the list of partial correlations at each step in Appendix A it is clear that variable 6 is a better predictor in most situations and therefore would be expected to do better in cross validation. The best 5, 6, and 7 predictor models are then those with asterisks by the R^2 values in Table 3. The choice between these models could be done after cross-validation and consideration of other criteria not discussed in this paper.

Table 3

Candidates for "Best" Model

Number of Predictors	Variables in the Model	R^2
5	2 5 9 12 13	.269878*
6	2 3 5 12 13 15	.299565
6	2 5 6 12 13 15	.296645*
7	2 5 6 9 12 13 15	.347206*
7	2 3 5 9 12 13 15	.346599

* "best" models

12

References

Berenson, M. L., et al. (1983). Intermediate statistical methods and applications. Englewood Cliffs, NJ: Prentice-Hall

Chatterjee, S. & Price, B. (1977). Regression Analysis by example. New York: John Wiley & Sons.

Cohen, J. & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Davidson, B. M. (1988). The case against using stepwise research methods. Paper presented at the Annual Meeting of the Mid-South Educational Research Association.

Draper, N. R. & Smith, H. (1981). Applied regression analysis, 2nd edition. New York: John Wiley & Sons.

Freund, R. J. & Minton, P. D. (1979). Regression methods. New York: Marcel Dekker.

Gunst, R. F. & Mason, R. L. (1980). Regression analysis and its application. New York: Marcel Dekker.

Huberty, C. J. (1989). Problems with stepwise methods--better alternatives. In Thompson, B. (Ed.), Advances in Social Science Methodology, vol 1, (pp. 43-70), Greenwich, CT: JAI Press.

Kleinbaum, D. G. & Kupper, L. L. (1988). Applied regression analysis and other multivariable methods (2nd Edition). Boston, MA: PWS-KENT.

Morrison, D. F. (1983). Applied linear statistical methods. Englewood Cliffs, NJ: Prentice-Hall.

Myers, R. H. (1986). Classical and modern regression with applications. Boston, MA: Duxbury.

Neter, J. et al. (1983). Applied linear regression models. Homewood, IL: Richard D. Irwin.

Pedhazur, E. J. (1982). Multiple regression in behavioral research, (2nd Edition). New York: Holt Rinehart and Winston.

Thayer, J. D. (1986). Testing different model building procedures using multiple regression. Multiple Linear Regression Viewpoints, 16, 37-52.

Thompson, B. (1989). Why won't stepwise methods die? Measurement and Evaluation in Counseling and Development, 21, 146-148.

Wilkinson, L. (1984). SYSTAT. Evanston, IL: SYSTAT.

Wittink, D. R. (1988). The application of regression analysis. Needham Heights, MA: Allyn and Bacon.

Younger, M. S. (1979). A handbook for linear regression. North Scituate, MA: Duxbury Press.

Appendix A

Forward/Backward Stepwise Results

```
BMDP2R - STEPWISE REGRESSION

PROGRAM INSTRUCTIONS

/input    file='a6'.
          format=free.
          variables=15.
/regress  dependent=1.
          independent=2 to 15.
          enter =0.01,500.
          remove=0.00,499.
/print    part.
/end

NUMBER OF CASES READ. . . . . .              50
```

VARIABLE NO.	NAME	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SKEWNESS	KURTOSIS	SMALLEST VALUE	LARGEST VALUE	SMALLEST STD SCORE	LARGEST STD SCORE
1	X(1)	0.3162	0.0482	0.152482	0.5209	-0.3210	0.2210	0.4270	-1.9745	2.2981
2	X(2)	177.9060	6.7031	0.037678	0.0684	0.0559	163.1000	193.5000	-2.2088	2.3264
3	X(3)	78.3524	11.4684	0.146369	0.3066	-0.7693	54.2800	102.2600	-2.0990	2.0847
4	X(4)	40.9520	1.5861	0.038730	-0.2923	-0.3913	37.2000	43.8000	-2.3656	1.7956
5	X(5)	28.1060	1.4372	0.051134	0.0975	0.1848	24.2000	31.6000	-2.7178	2.4312
6	X(6)	90.6200	5.9709	0.065890	0.2624	-1.0669	80.5000	101.2000	-1.6949	1.7719
7	X(7)	16.1300	6.1019	0.378293	0.0703	-0.1614	3.5000	32.0000	-2.0699	2.6008
8	X(8)	73.0800	12.9107	0.176666	0.4533	-0.5277	50.0000	104.0000	-1.7877	2.3949
9	X(9)	74.6000	8.1140	0.108767	-0.2439	0.9651	48.0000	90.0000	-3.2783	1.8980
10	X(10)	6.2800	4.3801	0.697469	0.8519	0.4282	0.0000	20.0000	-1.4338	3.1323
11	X(11)	193.3400	25.4547	0.131658	0.7104	0.7753	146.0000	272.0000	-1.8598	3.0902
12	X(12)	114.9400	10.3990	0.090473	0.6146	1.4698	91.0000	147.0000	-2.3021	3.0830
13	X(13)	5.4800	0.3637	0.066360	-1.1897	3.7250	4.0000	6.0000	-4.0698	1.4299
14	X(14)	4.1740	0.5642	0.135163	0.1375	-0.2559	3.0000	5.5000	-2.0809	2.3503
15	X(15)	13.7680	7.3521	0.533998	0.6392	-0.2236	2.0100	32.6300	-1.5993	2.5655

CORRELATIONS

		X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	X(9)	X(10)	X(11)	X(12)	X(13)	X(14)
		2	3	4	5	6	7	8	9	10	11	12	13	14
X(2)	2	1.000												
X(3)	3	0.635	1.000											
X(4)	4	0.654	0.666	1.000										
X(5)	5	0.586	0.647	0.582	1.000									
X(6)	6	0.426	0.889	0.554	0.522	1.000								
X(7)	7	0.223	0.554	0.205	0.207	0.398	1.000							
X(8)	8	-0.182	-0.264	-0.168	-0.323	-0.246	-0.006	1.000						
X(9)	9	-0.187	-0.052	-0.145	0.148	-0.008	0.049	0.234	1.000					
X(10)	10	-0.276	-0.576	-0.274	-0.158	-0.454	-0.670	0.155	0.054	1.000				
X(11)	11	0.588	0.450	0.368	0.354	0.347	0.207	-0.064	-0.165	-0.358	1.000			
X(12)	12	-0.151	-0.118	-0.042	-0.247	-0.065	0.244	0.528	0.155	-0.101	0.101	1.000		
X(13)	13	0.216	-0.053	0.202	0.041	-0.162	-0.208	-0.300	-0.321	0.324	-0.031	-0.473	1.000	
X(14)	14	-0.189	-0.368	-0.245	-0.229	-0.328	-0.336	0.007	0.134	0.213	-0.315	-0.018	-0.022	1.000
X(15)	15	0.364	0.810	0.329	0.414	0.727	0.843	-0.093	.0.047	-0.689	0.301	0.098	-0.205	-0.405
X(1)	1	0.222	0.056	-0.094	-0.056	-0.032	0.132	0.163	0.147	-0.158	0.160	-0.076	-0.149	-0.053

		X(15)	X(1)
		15	1
X(15)	15	1.000	
X(1)	1	0.165	1.000

Variable 2 had the highest zero-order correlations with x(1), the dependent variable. Variables 7-11, 13 and 15 all had similar values.

```
STEPPING ALGORITHM. . . . . .F
DEPENDENT VARIABLE. . . . . .      1 X(1)
MINIMUM ACCEPTABLE F TO ENTER . . . . . .  0.010, 500.000
MAXIMUM ACCEPTABLE F TO REMOVE. . . . . .  0.000, 499.000
MINIMUM ACCEPTABLE TOLERANCE. . . . . .0.01000
```

STEP NO. 2

VARIABLE ENTERED 4 X(4)

MULTIPLE R 0.3867
MULTIPLE R-SQUARE 0.1495
ADJUSTED R-SQUARE 0.1133

STD. ERROR OF EST. 0.0454

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.17033600E-01	2	0.8516802E-02	4.13
RESIDUAL	0.96874410E-01	47	0.2061158E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.20221)					
X(2)	2	0.00357	0.0013	0.496	0.57190	7.78	1
X(4)	4	-0.01272	0.0054	-0.418	0.57190	5.53	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	0.03034	0.48713	0.04	1
X(5)	5	-0.14554	0.58742	1.00	1
X(6)	6	-0.01457	0.68557	0.01	1
X(7)	7	0.11971	0.94419	0.67	1
X(8)	8	0.20199	0.96277	1.96	1
X(9)	9	0.19794	0.96429	1.88	1
X(10)	10	-0.15485	0.90860	1.13	1
X(11)	11	0.02914	0.65395	0.04	1
X(12)	12	-0.02049	0.97184	0.02	1
X(13)	13	-0.19154	0.94707	1.75	1
X(14)	14	-0.06840	0.93861	0.22	1
X(15)	15	0.14331	0.85272	0.96	1

This model was the model selected by forward stepwise using an F-to-enter of 2.00.

STEP NO. 21

VARIABLE REMOVED 7 X(7)

MULTIPLE R 0.5892
MULTIPLE R-SQUARE 0.3472
ADJUSTED R-SQUARE 0.2384

STD. ERROR OF EST. 0.0421

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.39549630E-01	7	0.5649948E-02	3.19
RESIDUAL	0.74358380E-01	42	0.1770438E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.57824)					
X(2)	2	0.00423	0.0012	0.588	0.53012	11.78	1
X(5)	5	-0.01570	0.0061	-0.468	0.47030	6.62	1
X(6)	6	-0.00318	0.0016	-0.393	0.38895	3.87	1
X(9)	9	0.15129E-02	0.8388E-03	0.255	0.77996	3.25	1
X(12)	12	-0.16762E-02	0.6988E-03	-0.362	0.68420	5.75	1
X(13)	13	-0.04381	0.0208	-0.330	0.63093	4.43	1
X(15)	15	0.00253	0.0012	0.386	0.43940	4.21	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.11216	0.07232	0.52	1
X(4)	4	-0.05438	0.36713	0.12	1
X(7)	7	-0.18517	0.18059	1.46	1
X(8)	8	0.14242	0.61554	0.85	1
X(10)	10	0.14081	0.42251	0.83	1
X(11)	11	0.10966	0.58982	0.50	1
X(14)	14	-0.09643	0.79427	0.38	1

This is the model that was selected by backward stepwise using an F-to-enter of 2.00.

STEP NO. 22

VARIABLE REMOVED 9 X(9)

MULTIPLE R 0.5447
MULTIPLE R-SQUARE 0.2966
ADJUSTED R-SQUARE 0.1985

STD. ERROR OF EST. 0.0432

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.33790250E-01	6	0.5631709E-02	3.02
RESIDUAL	0.80117760E-01	43	0.1863204E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.75673)					
X(2)	2	0.00364	0.0012	0.506	0.56993	8.92	1
X(5)	5	-0.01199	0.0059	-0.357	0.53053	4.14	1
X(6)	6	-0.00352	0.0016	-0.436	0.39451	4.59	1
X(12)	12	-0.15808E-02	0.7148E-03	-0.341	0.68815	4.89	1
X(13)	13	-0.05222	0.0208	-0.394	0.66426	6.30	1
X(15)	15	0.00261	0.0013	0.398	0.43996	4.26	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.13162	0.07290	0.74	1
X(4)	4	-0.08029	0.37123	0.27	1
X(7)	7	-0.18857	0.18085	1.55	1
X(8)	8	0.19456	0.64908	1.65	1
X(9)	9	0.26812	0.77996	3.25	1
X(10)	10	0.17102	0.43055	1.27	1
X(11)	11	0.06474	0.60307	0.18	1
X(14)	14	-0.04708	0.81680	0.09	1

This is the 6 predictor model that would have been selected by backward stepwise

STEP NO. 23

VARIABLE REMOVED 5 X(5)

MULTIPLE R 0.4784
MULTIPLE R-SQUARE 0.2289
ADJUSTED R-SQUARE 0.1412

STD. ERROR OF EST. 0.0447

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.26069730E-01	5	0.5213945E-02	2.61
RESIDUAL	0.87838280E-01	44	0.1996325E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.62171)					
X(2)	2	0.00250	0.0011	0.348	0.72136	4.99	1
X(6)	6	-0.00425	0.0017	-0.526	0.41403	6.54	1
X(12)	12	-0.12176E-02	0.7165E-03	-0.263	0.73387	2.89	1
X(13)	13	-0.04737	0.0214	-0.357	0.67307	4.90	1
X(15)	15	0.00245	0.0013	0.373	0.44178	3.51	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.19981	0.07841	1.79	1
X(4)	4	-0.16324	0.40909	1.18	1
X(5)	5	-0.29647	0.53053	4.14	1
X(7)	7	-0.16202	0.18150	1.16	1
X(8)	8	0.22815	0.66399	2.36	1
X(9)	9	0.14120	0.87984	0.87	1
X(10)	10	0.07498	0.46620	0.24	1
X(11)	11	0.05971	0.60311	0.15	1
X(14)	14	-0.02641	0.81998	0.03	1

This is the 5 predictor model that would have been selected by backward stepwise.

16

STEP NO. 24

VARIABLE REMOVED 12 X(12)

MULTIPLE R	0.4222
MULTIPLE R-SQUARE	0.1783
ADJUSTED R-SQUARE	0.1052
STD. ERROR OF EST.	0.0456

ANALYSIS OF VARIANCE

	SUM OF SQUARES	DF	MEAN SQUARE	F RATIO
REGRESSION	0.20304410E-01	4	0.5076102E-02	2.44
RESIDUAL	0.93603600E-01	45	0.2080080E-02	

VARIABLES IN EQUATION FOR X(1)

VARIABLE		COEFFICIENT	STD. ERROR OF COEFF	STD REG COEFF	TOLERANCE	F TO REMOVE	LEVEL
(Y-INTERCEPT		0.33832)					
X(2)	2	0.00250	0.0011	0.348	0.72136	4.78	1
X(6)	6	-0.00361	0.0017	-0.447	0.43646	4.77	1
X(13)	13	-0.03077	0.0194	-0.232	0.85029	2.51	1
X(15)	15	0.00207	0.0013	0.315	0.45494	2.48	1

VARIABLES NOT IN EQUATION

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER	LEVEL
X(3)	3	-0.13532	0.08226	0.82	1
X(4)	4	-0.21748	0.43907	2.18	1
X(5)	5	-0.21616	0.56578	2.16	1
X(7)	7	-0.20628	0.19016	1.96	1
X(8)	8	0.09303	0.81411	0.38	1
X(9)	9	0.14000	0.87999	0.88	1
X(10)	10	0.04373	0.47239	0.08	1
X(11)	11	0.01118	0.62425	0.01	1
X(12)	12	-0.24818	0.73387	2.89	1
X(14)	14	-0.01307	0.82206	0.01	1

This is the 4 predictor model that would have been selected by backward stepwise.

PARTIAL CORRELATIONS

STEP	VARIABLES	2 X(2)	3 X(3)	4 X(4)	5 X(5)	6 X(6)	7 X(7)	8 X(8)	9 X(9)	10 X(10)	11 X(11)
0		0.2223	0.0562	-0.0938	-0.0559	-0.0318	0.1324	0.1631	0.1473	-0.1585	0.1595
1		0.2223*	-0.1129	-0.3245	-0.2356	-0.1434	0.0871	0.2122	0.1971	-0.1036	0.0365
2		0.3768*	0.0303	-0.3245*	-0.1455	-0.0146	0.1197	0.2020	0.1979	-0.1548	0.0291
3		0.3975*	0.0704	-0.3183*	-0.0974	0.0223	0.1138	0.2020*	0.1632	-0.1804	0.0192
4		0.3828*	-0.0310	-0.3397*	-0.0846	-0.0460	-0.0057	0.2218*	0.1616	-0.1804*	-0.0300
5		0.3583*	-0.0544	-0.3269*	-0.1081	-0.0496	0.0388	0.2934*	0.1705	-0.2240*	0.0111
6		0.3754*	-0.0866	-0.3278*	-0.2068	-0.0762	0.0228	0.2693*	0.1705*	-0.2244*	0.0286
7		0.4204*	-0.0064	-0.2414*	-0.2068*	-0.0243	0.0658	0.2137*	0.2442*	-0.2151*	0.0606
8		0.4444*	-0.0245	-0.1903*	-0.2474*	-0.0717	0.1104	0.1864*	0.2220*	-0.1155*	0.0611
9		0.4462*	-0.2154	-0.1743*	-0.2824*	-0.2324	0.0115	0.1711*	0.2241*	0.0083*	0.0822
10		0.4536*	-0.0807	-0.0300*	-0.3116*	-0.2324*	-0.1459	0.1109*	0.2243*	0.0866*	0.1410
11		0.4489*	-0.0721	0.0063*	-0.3287*	-0.2721*	-0.1459*	0.0859*	0.2320*	0.0755*	0.1419
12		0.3634*	-0.0794	0.0392*	-0.3500*	-0.2936*	-0.1468*	0.0745*	0.2516*	0.1154*	0.1419*
13		0.3671*	-0.0794*	0.0607*	-0.3304*	-0.1897*	-0.1419*	0.0779*	0.2497*	0.1013*	0.1457*
14		0.3688*	-0.0785*	0.0549*	-0.3321*	-0.1896*	-0.1440*	0.0733*	0.2537*	0.0942*	0.1279*
15		0.3671*	-0.0794*	0.0607*	-0.3304*	-0.1897*	-0.1419*	0.0779*	0.2497*	0.1013*	0.1457*
16		0.3842*	-0.0645*	0.0607	-0.3326*	-0.1805*	-0.1323*	0.0930*	0.2442*	0.0867*	0.1354*
17		0.3951*	-0.0645	0.0392	-0.3731*	-0.3283*	-0.1417*	0.0862*	0.2487*	0.1088*	0.1367*
18		0.4118*	-0.0542	0.0586	-0.4059*	-0.3506*	-0.1521*	0.0862	0.2683*	0.1351*	0.1388*
19		0.3992*	-0.0872	0.0097	-0.3872*	-0.3470*	-0.1850*	0.1179	0.2791*	0.1351	0.1093*
20		0.4770*	-0.0818	-0.0050	-0.3809*	-0.3403*	-0.1852*	0.1135	0.2658*	0.1045	0.1093
21		0.4680*	-0.1122	-0.0544	-0.3691*	-0.2906*	-0.1852	0.1424	0.2681*	0.1408	0.1097
22		0.4146*	-0.1316	-0.0803	-0.2965*	-0.3105*	-0.1886	0.1946	0.2681	0.1710	0.0647
23		0.3191*	-0.1998	-0.1632	-0.2965	-0.3598*	-0.1620	0.2281	0.1412	0.0750	0.0597
24		0.3099*	-0.1353	-0.2175	-0.2162	-0.3097*	-0.2063	0.0930	0.1400	0.0437	0.0112
25		0.3321*	0.0332	-0.2502	-0.2058	-0.2167*	0.1024	0.1099	0.1484	-0.0946	0.0125
26		0.2636*	-0.1743	-0.3154	-0.2652	-0.2167	0.0333	0.1657	0.1459	-0.0212	-0.0051
27		0.2223*	-0.1129	-0.3245	-0.2356	-0.1434	0.0871	0.2122	0.1971	-0.1036	0.0365
		Good Always		Good Small Not Alone	Low Alone	Good Only With X(15)					

PARTIAL CORRELATIONS

STEP	VARIABLES	12 X(12)	13 X(13)	14 X(14)	15 X(15)
0		-0.0756	-0.1493	-0.0525	0.1650
1		-0.0437	-0.2072	-0.0110	0.0925
2		-0.0205	-0.1915	-0.0684	0.1433
3		-0.1511	-0.1458	-0.0618	0.1508
4		-0.2016	-0.0700	-0.0343	0.0443
5		-0.2016*	-0.1488	-0.0312	0.0587
6		-0.2087*	-0.1071	-0.0517	0.0320
7		-0.2363*	-0.1746	-0.0978	0.1194
8		-0.2836*	-0.1746*	-0.1150	0.1444
9		-0.3028*	-0.1922*	-0.0745	0.1444*
10		-0.3513*	-0.2726*	-0.0776	0.2626*
11		-0.3280*	-0.2823*	-0.0848	0.2718*
12		-0.3523*	-0.2970*	-0.0481	0.2904*
13		-0.3599*	-0.2973*	-0.0467	0.3000*
14		-0.3554*	-0.2983*	-0.0467*	0.2910*
15		-0.3599*	-0.2973*	-0.0467	0.3000*
16		-0.3631*	-0.3008*	-0.0534	0.3119*
17		-0.3581*	-0.3122*	-0.0524	0.3059*
18		-0.3520*	-0.3308*	-0.0625	0.3274*
19		-0.3298*	-0.3048*	-0.0803	0.3080*
20		-0.3156*	-0.3138*	-0.1028	0.3064*
21		-0.3471*	-0.3090*	-0.0964	0.3020*
22		-0.3196*	-0.3575*	-0.0471	0.3004*
23		-0.2482*	-0.3166*	-0.0264	0.2717*
24		-0.2482	-0.2298*	-0.0131	0.2285*
25		-0.1992	-0.2622*	-0.0736	0.2285
26		-0.1587	-0.2072*	-0.0072	0.0298
27		-0.0437	-0.2072	-0.0110	0.0925
		Good Large	Good Large		Good Large

SUMMARY TABLE

STEP NO.	VARIABLE ENTERED	VARIABLE REMOVED	MULTIPLE R	MULTIPLE RSQ	CHANGE IN RSQ	F TO ENTER	F TO REMOVE	NO.OF VAR. INCLUDED
1	2 X(2)		0.2223	0.0494	0.0494	2.50		1
2	4 X(4)		0.3867	0.1495	0.1001	5.53		2
3	8 X(8)		0.4292	0.1842	0.0347	1.96		3
4	10 X(10)		0.4591	0.2108	0.0265	1.51		4
5	12 X(12)		0.4928	0.2428	0.0321	1.86		5
6	9 X(9)		0.5146	0.2649	0.0220	1.29		6
7	5 X(5)		0.5443	0.2963	0.0314	1.88		7
8	13 X(13)		0.5637	0.3178	0.0215	1.29		8
9	15 X(15)		0.5762	0.3320	0.0142	0.85		9
10	6 X(6)		0.6067	0.3681	0.0361	2.23		10
11	7 X(7)		0.6177	0.3815	0.0135	0.83		11
12	11 X(11)		0.6277	0.3940	0.0125	0.76		12
13	3 X(3)		0.6307	0.3978	0.0038	0.23		13
14	14 X(14)		0.6318	0.3991	0.0013	0.08		14
15		14 X(14)	0.6307	0.3978	-0.0013		0.08	13
16		4 X(4)	0.6289	0.3956	-0.0022		0.13	12
17		3 X(3)	0.6269	0.3930	-0.0025		0.15	11
18		8 X(8)	0.6233	0.3885	-0.0045		0.28	10
19		10 X(10)	0.6141	0.3771	-0.0114		0.73	9
20		11 X(11)	0.6079	0.3696	-0.0075		0.48	8
21		7 X(7)	0.5892	0.3472	-0.0224		1.46	7
22		9 X(9)	0.5447	0.2966	-0.0506		3.25	6
23		5 X(5)	0.4784	0.2289	-0.0678		4.14	5
24		12 X(12)	0.4222	0.1783	-0.0506		2.89	4
25		15 X(15)	0.3647	0.1330	-0.0453		2.48	3
26		6 X(6)	0.3004	0.0903	-0.0427		2.27	2
27		13 X(13)	0.2223	0.0494	-0.0408		2.11	1

Appendix B

All Possible Subsets Output

```
BMDP9R - ALL POSSIBLE SUBSETS REGRESSION

PROGRAM INSTRUCTIONS

/input    file='a6'.
          format=free.
          variables=15.
/regress  dependent=1.
          independent=2 to 15.
/end

DEPENDENT VARIABLE. . . . . .        1 X(1)
NUMBER OF 'BEST' REGRESSIONS REPORTED . . . . .      5
SELECTION CRITERION . . . . . CP

NUMBER OF CASES READ. . . . . .      50
```

FOR EACH SUBSET SELECTED BY YOUR CRITERION, THE R-SQUARED, ADJUSTED R-SQUARED, MALLOWS' CP, AND THE VARIABLE NAMES ARE PRINTED. THE REGRESSION COEFFICIENTS AND T-STATISTICS ARE PRINTED TO THE RIGHT OF THE VARIABLE NAMES. MANY OTHER SUBSETS MAY ALSO BE REPORTED THAT ARE NOT ACCOMPANIED BY REGRESSION COEFFICIENTS AND T-STATISTICS. SOME OF THESE SUBSETS MAY BE QUITE GOOD ALTHOUGH THEY ARE NOT NECESSARILY BETTER THAN ANY SUBSET THAT HAS NOT BEEN PRINTED.

SUBSETS WITH 2 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP					
0.149539	0.113349	5.54	X(2)	X(4)			Forward Stepwise 2 Predictor (BEST) Model
0.102208	0.064004	8.29	X(2)	X(5)			
0.092249	0.053621	8.87	X(2)	X(8)			
0.090255	0.051543	8.99	X(2)	X(13)			Backward Stepwise 2 Predictor Model

SUBSETS WITH 4 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP					
0.227830	0.159193	4.98	X(2)	X(4)	X(5)	X(9)	
0.210777	0.140623	5.97	X(2)	X(4)	X(8)	X(10)	Forward Stepwise 4 Predictor Model
							Backward Stepwise 4 Predictor Model Not Listed In Best Ten 4 Predictor Models

SUBSETS WITH 5 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP	VARIABLE	COEFFICIENT	T-STATISTIC			
0.269878	0.186910	4.53	2 X(2)	0.00425862	3.43			
			5 X(5)	-0.0169603	-2.88			
			9 X(9)	0.00166493	1.94			
			12 X(12)	-0.00139223	-1.96			
			13 X(13)	-0.0408527	-1.98			
			INTERCEPT	0.294943				
0.252855	0.167953	5.52	X(2)	X(3)	X(12)	X(13)	X(15)	
0.247058	0.161496	5.86	X(2)	X(4)	X(5)	X(9)	X(13)	
0.244840	0.159026	5.99	X(2)	X(5)	X(8)	X(12)	X(13)	
0.242845	0.156804	6.10	X(2)	X(4)	X(8)	X(10)	X(12)	Forward Stepwise 4 Predictor Model
								Backward Stepwise 5 Predictor Model Not Listed In Best Ten 5 Predictor Models

SUBSETS WITH 6 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP	VARIABLE	COEFFICIENT	T-STATISTIC
0.299565	0.201830	4.80	2 X(2)	0.00469572	3.44
			3 X(3)	-0.00285065	-2.19
			5 X(5)	-0.0103034	-1.69
			12 X(12)	-0.00168170	-2.33
			13 X(13)	-0.0488981	-2.38
			15 X(15)	0.00369285	2.30
			INTERCEPT	0.404160	

0.296645	0.198502	4.97	X(2)	X(5)	X(6)	X(12)	X(13)	X(15)	Backward Stepwise 6 Predictor Model
0.291666	0.192828	5.26	X(2)	X(3)	X(8)	X(12)	X(13)	X(15)	
0.290071	0.191011	5.35	X(2)	X(4)	X(5)	X(9)	X(12)	X(13)	
0.288310	0.189004	5.45	X(2)	X(5)	X(8)	X(9)	X(12)	X(13)	
0.287008	0.187520	5.53	X(2)	X(5)	X(9)	X(12)	X(13)	X(15)	
0.285008	0.185242	5.65	X(2)	X(3)	X(5)	X(9)	X(12)	X(15)	
0.282375	0.182242	5.80	X(2)	X(5)	X(9)	X(12)	X(13)	X(14)	
0.282028	0.181846	5.82	X(2)	X(5)	X(7)	X(9)	X(12)	X(13)	
0.281706	0.181479	5.84	X(2)	X(5)	X(6)	X(9)	X(12)	X(13)	

Forward Stepwise 6 Predictor Model
Not Listed In Best Ten 6 Predictor Models

SUBSETS WITH 7 VARIABLES

R-SQUARED	ADJUSTED R-SQUARED	CP	VARIABLE	COEFFICIENT	T-STATISTIC
0.347206	0.238407	4.02	2 X(2)	0.00422728	3.43
			5 X(5)	-0.0156974	-2.57
			6 X(6)	-0.00317676	-1.97
			9 X(9)	0.00151294	1.80
			12 X(12)	-0.00167624	-2.40
			13 X(13)	-0.0438120	-2.11
			15 X(15)	0.00253197	2.05
			INTERCEPT	0.578242	

Backward Stepwise 7 Predictor (BEST) Model

0.346599	0.237699	4.06	VARIABLE	COEFFICIENT	T-STATISTIC
			2 X(2)	0.00513442	3.78
			3 X(3)	-0.00251951	-1.96
			5 X(5)	-0.0141462	-2.23
			9 X(9)	0.00146501	1.74
			12 X(12)	-0.00175720	-2.49
			13 X(13)	-0.0409863	-1.99
			15 X(15)	0.00345443	2.19
			INTERCEPT	0.267485	

Forward Stepwise 7 Predictor Model
Not Listed In Best Ten 7 Predictor Models

Appendix C

Model 1 -- All Possible Subsets Best 7 Predictor Model & Backward Stepwise 7 Predictor Model

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.02
SQUARED MULTIPLE CORRELATION	0.34721
MULTIPLE CORRELATION	0.58924
ADJUSTED SQUARED MULT. CORR.	0.23841
RESIDUAL MEAN SQUARE	0.001770
STANDARD ERROR OF EST.	0.042077
F-STATISTIC	3.19
NUMERATOR DEGREES OF FREEDOM	7
DENOMINATOR DEGREES OF FREEDOM	42
SIGNIFICANCE (TAIL PROB.)	0.0084

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ	
	INTERCEPT	0.578242	0.275920	11.993	2.10	0.042			
2	X(2)	0.00422728	0.00123163	0.588	3.43	0.001	0.530117	0.18310	
5	X(5)	-0.0156974	0.00609880	-0.468	-2.57	0.014	0.470301	0.10297	2nd Highest Contribution to R-SQ
6	X(6)	-0.00317676	0.00161419	-0.393	-1.97	0.056	0.388951	0.06020	
9	X(9)	0.00151294	0.000838828	0.255	1.80	0.078	0.779956	0.05056	
12	X(12)	-0.00167624	0.000698809	-0.362	-2.40	0.021	0.684202	0.08943	
13	X(13)	-0.0438120	0.0208095	-0.330	-2.11	0.041	0.630935	0.06890	
15	X(15)	0.00253197	0.00123339	0.386	2.05	0.046	0.439403	0.06550	

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

Model 2 -- All Possible Subsets Best 6 Predictor Model

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	0.29956
MULTIPLE CORRELATION	0.54733
ADJUSTED SQUARED MULT. CORR.	0.20183
RESIDUAL MEAN SQUARE	0.001855
STANDARD ERROR OF EST.	0.043075
F-STATISTIC	3.07
NUMERATOR DEGREES OF FREEDOM	6
DENOMINATOR DEGREES OF FREEDOM	43
SIGNIFICANCE (TAIL PROB.)	0.0138

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ	
	INTERCEPT	0.404160	0.248288	8.383	1.63	0.111			
2	X(2)	0.00469572	0.00136419	0.653	3.44	0.001	0.452848	0.19300	
3	X(3)	-0.00285065	0.00130277	-0.678	-2.19	0.034	0.169636	0.07799	
5	X(5)	-0.0103034	0.00608455	-0.307	-1.69	0.098	0.495200	0.04671	Lowest Contribution to R-SQ
12	X(12)	-0.00168170	0.000721680	-0.363	-2.33	0.025	0.672333	0.08845	
13	X(13)	-0.0488981	0.0205573	-0.369	-2.38	0.022	0.677556	0.09216	
15	X(15)	0.00369285	0.00160697	0.563	2.30	0.026	0.271283	0.08602	

Model 3 -- Backward Stepwise 6 Predictor Model

SQUARED MULTIPLE CORRELATION	0.29664
MULTIPLE CORRELATION	0.54465
ADJUSTED SQUARED MULT. CORR.	0.19850
RESIDUAL MEAN SQUARE	0.001863
STANDARD ERROR OF EST.	0.043165
F-STATISTIC	3.02
NUMERATOR DEGREES OF FREEDOM	6
DENOMINATOR DEGREES OF FREEDOM	43
SIGNIFICANCE (TAIL PROB.)	0.0148

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
	INTERCEPT	0.756732	0.264225	15.695	2.86	0.006			
2	X(2)	0.00364014	0.00121855	0.506	2.99	0.005	0.569933	0.14597	
5	X(5)	-0.0119911	0.00589071	-0.357	-2.04	0.048	0.530530	0.06778	Lowest Contribution to R-SQ
6	X(6)	-0.00352246	0.00164422	-0.436	-2.14	0.038	0.394513	0.07507	
12	X(12)	-0.00158083	0.000714826	-0.341	-2.21	0.032	0.688146	0.08000	
13	X(13)	-0.0522187	0.0208053	-0.394	-2.51	0.016	0.664260	0.10304	
15	X(15)	0.00261129	0.00126449	0.398	2.07	0.045	0.439962	0.06976	

Model 4 -- All Possible Subsets Best 5 Predictor Model

SQUARED MULTIPLE CORRELATION	0.26988
MULTIPLE CORRELATION	0.51950
ADJUSTED SQUARED MULT. CORR.	0.18691
RESIDUAL MEAN SQUARE	0.001890
STANDARD ERROR OF EST.	0.043476
F-STATISTIC	3.25
NUMERATOR DEGREES OF FREEDOM	5
DENOMINATOR DEGREES OF FREEDOM	44
SIGNIFICANCE (TAIL PROB.)	0.0138

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ	
	INTERCEPT	0.294943	0.253049	6.117	1.17	0.250			
2	X(2)	0.00425862	0.00124108	0.592	3.43	0.001	0.557368	0.19538	
5	X(5)	-0.0169603	0.00589221	-0.506	-2.88	0.006	0.537929	0.13748	2nd Highest Contribution to R-SQ
9	X(9)	0.00166493	0.000859501	0.280	1.94	0.059	0.793122	0.06226	
12	X(12)	-0.00139223	0.000709751	-0.300	-1.96	0.056	0.708118	0.06385	
13	X(13)	-0.0408527	0.0205961	-0.308	-1.98	0.054	0.687627	0.06529	